ECO 519. Handout on Quantile Regression

1 Setup

The classic reference is Koenker and Basset (1978). We start by noting the result that suppose a random variable z_i has distribution given by $F_z(\cdot)$. Take $\tau \in (0,1)$, then the solution to the problem

$$\underset{b}{\text{Minimize}}\ E\Big[\tau \cdot \big|z_i - b\big|1\!\!1\{z_i - b \geq 0\} + (1-\tau)\cdot \big|z_i - b\big|1\!\!1\{z_i - b < 0\}\Big]$$

is given by $b = F_z^{-1}(\tau)$, or the τ^{th} quantile of z_i. We can show this immediately if we assume that z_i has a continuous distribution:

$$E\Big[\tau\cdot\big|z_i - b\big|1\!\!1\{z_i - b \geq 0\} + (1-\tau)\cdot\big|z_i - b\big|1\!\!1\{z_i - b < 0\}\Big] = \tau\int_b^{\infty}(z-b)f_z(z)dz - (1-\tau)\int_{-\infty}^{b}(z-b)f_z(z)dz$$

and taking first order conditions with respect to b yields (via Leibniz' Rule) $F_z(b) = \tau$, or $b = F_z^{-1}(\tau)$. Now consider the linear model

$$y_i = \beta_0' x_i + u_i,$$

where $u_i \sim F_u(\cdot)$ is continuously distributed and independent of x_i. The τ^{th}-quantile of u_i is given by $F_u^{-1}(\tau)$ independently of x_i. Suppose from the onset that x_i includes a constant (i.e, an intercept). Consider now the problem

$$\underset{\beta}{\text{Minimize}}\ E\Big[\tau\cdot\big|y_i - x_i'\beta\big|1\!\!1\{y_i - x_i'\beta \geq 0\} + (1-\tau)\cdot\big|y_i - x_i'\beta\big|1\!\!1\{y_i - x_i'\beta < 0\}\Big|x_i\Big].$$

This is simply

$$\underset{\beta}{\text{Min}}\ E\Big[\tau\cdot\big|u_i - x_i'(\beta-\beta_0)\big|1\!\!1\{u_i - x_i'(\beta-\beta_0) \geq 0\} + (1-\tau)\cdot\big|u_i - x_i'(\beta-\beta_0)\big|1\!\!1\{u_i - x_i'(\beta-\beta_0) < 0\}\Big|x_i\Big].$$

As we saw above, the solution to this problem must be the conditional τ^{th}-quantile of u_i given x_i. That is, for all x_i we must have

$$x_i'(\beta-\beta_0) = F_u^{-1}(\tau) \ \Rightarrow x_i'\beta = x_i'\beta_0 + F_u^{-1}(\tau),$$

which (given the fact that x_i includes an intercept) is solved by using $\beta = \beta_0 + \boldsymbol{\iota}F_u^{-1}(\tau)$, where $\boldsymbol{\iota} = (1, 0, \ldots, 0)$. We denote this shortly by $\beta(\tau)$ and assume that it is the unique solution (think about the usual full-rank assumption). Note that if the τ^{th} quantile of u_i is zero, then $\beta(\tau) = \beta_0$ and most importantly, notice that $\beta(\tau)$ and β_0 differ only in the intercept. Notice that all the action takes place in the intercept.

2 Estimation

Using the analogy principle, we can estimate $\beta(\tau)$ by solving

$$\operatorname*{Min}_{\beta} \frac{1}{N}\sum_{i=1}^{N}\Big(\tau \cdot \big|y_i - x_i'\beta\big| \mathbb{1}\{y_i - x_i'\beta \geq 0\} + (1-\tau)\cdot \big|y_i - x_i'\beta\big|\mathbb{1}\{y_i - x_i'\beta < 0\}\Big),$$

which can be simplified to

$$\operatorname*{Min}_{\beta} \frac{1}{N}\sum_{i=1}^{N}\big|y_i - x_i'\beta\big|\cdot\big[\tau + (1-2\tau)\mathbb{1}\{y_i - x_i'\beta < 0\}\big]$$

After minor simplification, the left-partial derivative with respect to β_ℓ is easily given by

$$S_{N,\ell}(\beta) \equiv \frac{1}{N}\sum_{i=1}^{N} x_{i,\ell}\big(\mathbb{1}\{y_i < x_i'\beta\} - \tau\big).$$

Now remember the key argument we used when discussing Powell's Censored LAD. Since $\widehat{\beta}$ is minimizing the objective function, it must be case that for any $\delta > 0$:

$$S_{N,\ell}(\widehat{\beta}_\ell - \delta, \widehat{\beta}_{-\ell}) \leq S_{N,\ell}(\widehat{\beta}) \leq S_{N,\ell}(\widehat{\beta}_\ell + \delta, \widehat{\beta}_{-\ell}) \;\Rightarrow\; \big|S_{N,\ell}(\widehat{\beta})\big| \leq \big|S_{N,\ell}(\widehat{\beta}_\ell + \delta, \widehat{\beta}_{-\ell}) - S_{N,\ell}(\widehat{\beta}_\ell - \delta, \widehat{\beta}_{-\ell})\big|$$

and this holds for all $\ell = q, \ldots, K \equiv \dim(\beta)$. This implies that

$$\begin{aligned}\big|S_{N,\ell}(\widehat{\beta})\big| &\leq \frac{1}{N}\sum_{i=1}^{N}\Big|x_{i,\ell}\big(\mathbb{1}\{y_i < x_i'\widehat{\beta} + x_{i,\ell}\} - \mathbb{1}\{y_i < x_i'\widehat{\beta} - x_{i,\ell}\}\big)\Big| \\ &= \frac{1}{N}\sum_{i=1}^{N}\big|x_{i,\ell}\big|\cdot \mathbb{1}\{x_i'\widehat{\beta} - |x_{i,\ell}|\delta < y_i < x_i'\widehat{\beta} + |x_{i,\ell}|\delta\}\end{aligned}$$

Since this holds for any $\delta > 0$, letting $\delta \to 0$ we get

$$\sqrt{N}\big|S_{N,\ell}(\widehat{\beta})\big| \leq \operatorname*{Max}_{i=1,\ldots,N}\big|x_{i,\ell}\big|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbb{1}\{y_i = x_i'\widehat{\beta}\} = \operatorname*{Max}_{i=1,\ldots,N}\big|x_{i,\ell}\big|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbb{1}\{u_i = x_i'(\widehat{\beta} - \beta_0)\}$$

If we assume that $\big\|x_i\big\|$ is well-behaved in the sense for example that there exists a random variable z_i such that $\big\|x_i\big\| \leq z_i$ w.p.1, and $E\big|z_i\big| < \infty$, then $S_N(\beta)$ is easily uniformly bounded by $\frac{1}{N}\sum_{i=1}^{N} z_i$. This yields uniform convergence of $S_{N}(\beta)$ to $E\big[x_i\big(F_u(x_i'(\beta-\beta_0)-\tau)\big)\big]$, which is uniquely minimized at $\beta(\tau)$. The continuously-distributed nature of u_i and the fact that $\widehat{\beta} \xrightarrow{p} \beta(\tau)$ yields $\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \mathbb{1}\{u_i = x_i'(\widehat{\beta}-\beta_0)\} = o_p(1)$. Therefore

$$\underbrace{\operatorname*{Max}_{i=1,\ldots,N}\big|x_{i,\ell}\big|}_{=O_p(1)}\underbrace{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbb{1}\{u_i = x_i'(\widehat{\beta}-\beta_0)\}}_{=o_p(1)} = o_p(1)$$

and therefore $\sqrt{N}S_{N,\ell}(\widehat{\beta}) = o_p(1)$. To find the asymptotic distribution of $\sqrt{N}(\widehat{\beta}-\beta(\tau))$, the best and fastest thing to do is to analyze the empirical process:

$$\nu_{N}(\beta) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Big[x_i\big(\mathbb{1}\{y_i < x_i'\beta\}-\tau\big) - E\big[x_i\big(\mathbb{1}\{y_i < x_i'\beta\}-\tau\big)\big]\Big].$$

This is easily a vector-valued empirical process spanned by Type-I functions (Andrews, pp.2270). In addition, the empirical process $\nu_{N}(\beta)$ has envelope z_i, so if we assume $E[z_i^{2+\delta}] < \infty$ for some $\delta > 0$, then all the assumptions in Theorem 1 in Andrews. Therefore $\nu_{N}(\beta)$ is stochastically equicontinuous. Therefore, the asymptotic distribution of $\sqrt{N}(\widehat{\beta}-\beta(\tau))$ can be immediately derived as in Section 3.2 in Andrews.